

1 **Estimating fine-scale movement rates and habitat preferences using multiple**
2 **data sources**

3

4 Running header: Synthesis of high-resolution movement

5

6 James T. Thorson^{1,*}, Steven J. Barbeaux², Daniel R. Goethel³, Kelly A. Kearney^{4,5}, Ned Laman⁶,
7 Julie Nielsen⁷, Matt Siskey⁸, Kevin Siwicke³, Grant G. Thompson²

8

9 ¹ Habitat and Ecological Processes Research Program, Alaska Fisheries Science Center, NOAA.

10 ² Status of Stocks and Multispecies Assessments, Resource Ecology and Fisheries Management
11 Division, Alaska Fisheries Science Center, NOAA

12 ³ Marine Ecology and Stock Assessment, Auke Bay Laboratories, Alaska Fisheries Science
13 Center, NOAA

14 ⁴ Cooperative Institute for Climate, Ocean and Ecosystem Studies, University of Washington

15 ⁵ Resource Ecology and Ecosystem Modeling Program, Resource Ecology and Fisheries
16 Management Division, Alaska Fisheries Science Center, NOAA

17 ⁶ Groundfish Assessment Program, Resource Assessment and Conservation Division, Alaska
18 Fisheries Science Center, NOAA

19 ⁷ Kingfisher Marine Research, LLC

20 * Corresponding author: 7600 Sand Point Way NE, Seattle, WA 98115,
21 James.Thorson@noaa.gov

22

23 **Abstract:**

24 Fisheries scientists and managers must track rapid shifts in fish spatial distribution, mitigate
25 stakeholder conflict, and optimize survey designs, and these spatial shifts result in part from
26 animal movement. Information regarding animal movement can be obtained from selection
27 experiments, tagging studies, flux through movement gates (e.g., acoustic arrays), fishery catch-
28 per-unit effort (CPUE), resource surveys, and genetic/chemical markers. However, there are few
29 accessible approaches to combine these data types while accounting for spatially correlated
30 residual patterns. We therefore discuss a movement model involving diffusion (random
31 movement), taxis (movement towards preferred habitat), and advection (passive drift following
32 ocean currents). We specifically outline how these movement processes can be fitted to data
33 while discretizing space and time, and estimating nonlinear habitat preferences using
34 environmental layers as well as spatial process errors. Finally, we introduce an R package, ATM,
35 by fitting the model to bottom trawl survey, longline fishery, and tagging data for Pacific cod
36 (*Gadus macrocephalus*, Gadidae) in the Bering Sea during winter/summer seasons from 1982-
37 2019. Combining data types predicts an increasing proportion of cod residing in the northern
38 Bering Sea from 2013-2019, and estimates are informative in a recent stock assessment model.
39 We fit sensitivity analyses by dropping tag, survey, or fishery data, and this analysis shows that
40 tagging data are necessary to identify diffusion rates, while survey data are informative about
41 movement among biogeographic strata. This “hybrid” species distribution model can help
42 explain poleward movement, project distributions under future climate conditions, and evaluate
43 alternative tag-deployment scenarios to optimize tagging designs.

44 Keywords: advection; diffusion; habitat preference; instantaneous movement; Pacific cod; taxis

45	Table of contents
46	1. Introduction
47	2. Methods
48	a. Diffusion-advection-taxis movement process
49	b. Spatio-temporal model for spatial distribution
50	c. Metrics of resulting distribution shift
51	d. Parameter estimation
52	e. Case-study application
53	i. Investigating performance
54	ii. Sensitivity analyses
55	3. Results
56	4. Discussion
57	

58 **Introduction**

59 Understanding animal migrations and distributional shifts is central to science supporting
60 ocean management, including stock, ecosystem, climate, habitat, and protected species
61 assessments routinely conducted in the United States, Europe, and worldwide. For example,
62 climate-driven shifts in the spatial distribution of humpback whales led to increased
63 entanglements in the valuable Dungeness Crab (*Cancer magister*) fishery off California in 2014-
64 2016 (Santora et al., 2020), and decreasing wintertime sea ice is leading to increased overlap
65 between shipping routes and protected species migrations in the Arctic and adjacent oceans
66 (Hauser et al., 2018). Similarly, understanding movement and resulting distribution shifts is
67 necessary to evaluate the impact of localized habitat protections upon population productivity at
68 geographically distant locations that are linked via ontogenetic habitat shifts. For these and other
69 reasons, improved understanding of movement has been called a “Grand Challenge” for habitat
70 science (Thorson et al., 2021).

71 The past decade has seen an explosion of new and cost-effective technologies to study
72 individual movements of marine species (Lowerre-Barbieri et al., 2019), resulting in broad
73 deployment of many sampling methods including the following:

- 74 1. Conventional tags providing release and recovery location at known release and recovery
75 dates.
- 76 2. Electronic (satellite and/or archival) tags providing environmental measurements that can be
77 used to estimate animal location and behavior between release and recovery (Lam et al.,
78 2008).

- 79 3. Point-count data from resource surveys, providing a snapshot of habitat utilization resulting
80 from movement.
- 81 4. Point-count data from fishery operations, typically occurring over a wider seasonal interval
82 than surveys, where locational choices and gear deployment are not experimentally
83 controlled and therefore must be corrected for during analysis.
- 84 5. Movement gates (e.g., upward facing acoustics, acoustic telemetry arrays, environmental
85 DNA, acoustical moorings, and videocamera arrays) providing a high-frequency
86 measurement of aggregate flux (the product of local density and movement rates) and/or
87 average residence time at a given location (Shertzer & Bacheler, 2020).
- 88 6. Selection experiments (and similar process-research designs) specifying exogenous
89 treatments in laboratory or field conditions to provide robust inference about the causal
90 relationship between environmental features and movement decisions (e.g., Laurel et al.,
91 2007).
- 92 7. Chemical, genetic, and biological tracers, providing information on parentage (e.g., based on
93 genetic markers of sub-population structure), natal origin, or lifetime habitat utilization (e.g.,
94 based on stable isotope, trace element, or parasite markers of estuarine, nearshore, and
95 offshore habitat) (e.g., Spies et al., 2020).

96 These data types (Table 1) see varying usage across marine to freshwater systems. For example,
97 conventional tags are often analyzed in combination with fishery-dependent data for high-value
98 commercial tunas that are broadly distributed across tropical and temperate oceans (Fournier et
99 al., 1998; Lehodey et al., 2008), while surveys and movement gates are seeing increased use to
100 infer movement in coastal shelf systems for species that cannot reliably be captured and released
101 to deploy tags (O’Leary, Thorson, Ianelli, et al., 2020; Pinsky et al., 2013).

102 Given rapid declines in cost for new tagging and tracking technologies, researchers have
103 declared 2018-2028 the “Biologging decade” (Lowerre-Barbieri et al., 2019). However, rapid
104 deployment of new tracking technologies also raises analytical challenges, whereby new data
105 types must be integrated with existing knowledge and monitoring programs to provide a
106 synthetic picture of animal movements (Hays et al., 2019). There is an ongoing effort to
107 integrate all available data in its “rawest” form feasible throughout ecology, including (but not
108 limited to) marine stock assessment (Maunder & Punt, 2013), terrestrial monitoring programs
109 (Kéry & Schaub, 2021), life-history theory (Thorson, 2020), and evolutionary analysis (Lee &
110 Palci, 2015). Integrated models typically combine multiple data types to rectify deficiencies that
111 arise in individual data sets (e.g., combining spatially extensive occupancy data with replicated
112 counts at key monitoring sites) (Zipkin et al., 2017). Similarly, alternative movement
113 technologies provide complementary information where, for example, movement gates (eDNA
114 autosamplers and acoustics) can provide high temporal frequency at fixed locations, while
115 resource surveys provide spatially distributed information about resulting distribution and
116 density at seasonal or annual scales.

117 Spatially explicit integrated populations models (IPMs) have increasingly become useful
118 tools for identifying changes in population status and distribution to manage human impacts on
119 natural resources (Berger et al., 2017). Conventional stock-assessment models make implicit
120 assumptions about spatial processes, and spatial IPMs can improve these by letting data apply to
121 only the portion of the population from which it was collected. However, spatial IPMs require
122 specifying the fraction of abundance in each stratum that moves to other spatial strata (which we
123 call “movement fractions”); information to specifying these movement fractions remains limited
124 for spatial IPMs, especially for marine species where visual sampling of precise movement

125 tracks can be difficult to attain (Goethel et al., 2021). Because fisheries management often relies
126 on complex spatiotemporal policies (e.g., marine protected areas and area-specific quota
127 recommendations), there is increasing need for the development of stock assessment models that
128 are able to separately track abundance and age/size-structure across multiple spatial strata or
129 interconnected population components, including straying among population or management
130 units (Berger et al., 2017).

131 Although spatially stratified stock assessment models have been developed that can estimate
132 movement fractions among spatial strata by incorporating typical fishery data as well as
133 conventional, electronic, and natural tags (Hampton & Fournier, 2001; Methot & Wetzel, 2013),
134 there have been no previous models that simultaneously account for fine-scale spatial dynamics
135 and spatially correlated process errors when fitting data from tags. Instead, fine-scale spatial
136 IPMs such as SPM (Dunn et al., 2015) and SEAPODYM (Lehodey et al., 2008; Senina et al.,
137 2019) use habitat-preference functions to improve precision about local movement but lack
138 spatially correlated process errors, while other fine-scale IPMs include spatial process errors but
139 do not fit to tagging data (Cao et al., 2020; Kristensen et al., 2014). Additionally, no previous
140 study has proposed methods to incorporate the full array of data that are informative about
141 connectivity, movement, migration, and habitat preference in marine populations (listed in Table
142 1). As IPMs continue to integrate insight from the movement ecology paradigm (Nathan et al.,
143 2008), analysts require methods that can incorporate the full array of available spatial data,
144 incorporate information from habitat and oceanographic variables, and explicitly estimate
145 movement (Bruneel et al., 2018). This synthesis will improve spatial IPMs by informing or
146 directly specifying movement fractions and/or habitat preferences, thus reducing the number of
147 estimated parameters (Goethel et al., 2021).

148 As an alternative to fitting tags within IPMs, parallel research has developed a growing suite
149 of models (and associated estimation approaches) for individual movement at high spatial and
150 temporal resolution. This includes, e.g., a correlated random walk describing the animal location
151 as a latent variable across time (Johnson et al., 2008) or fast approximations that discretize space
152 and track animal residence within these discrete spatial cells while defining movement as a
153 continuous-time Markov chain, CTMC (Wikle, 2003; Hanks et al., 2015). These approaches can
154 then incorporate a habitat preference (a.k.a. resource selection) function to inform movement
155 (Preisler et al., 2013; Michelot et al., 2019), and the stationary distribution of this CTMC is then
156 a prediction of long term habitat utilization (Wilson et al., 2018). Despite these improvements,
157 there are few examples of CTMC or other tracking models that integrate point-count data from
158 surveys and fisheries, although again noting Lehodey et al. (2008) and Senina et al. (2019) as
159 exceptions.

160 In this study, we demonstrate the potential to integrate a broad range of information
161 regarding movement in marine species within a single model with high spatial resolution. The
162 approach decomposes an instantaneous movement rate into components representing diffusion,
163 advection, and taxis, and incorporates climate and habitat information to inform all three
164 processes. It then integrates these movement processes over seasonal and annual intervals to
165 identify path-dependent movement probabilities, and uses these probabilities to fit the movement
166 model to survey, fishery, and tagging data. Throughout, we specifically discuss how to extend
167 this framework to integrate information from movement gates, archival tags, and other
168 alternative technologies, while providing a publicly available R package to demonstrate the
169 analytical approach. In particular, we demonstrate this R package by fitting summer survey,
170 summer and winter fishery catch-and-effort, and conventional tagging data for Pacific cod

171 (*Gadus macrocephalus*) in the Bering Sea, using bottom temperature and bathymetry data to
172 interpret their seasonal habitat preferences. Finally, we demonstrate how results can be
173 integrated into an existing stock assessment model for Pacific cod, which has exhibited rapid
174 poleward distribution shifts in response to warming conditions over the past decade. We also
175 explore the sensitivity of model results to each fitted data set to determine the relative
176 importance of tags, survey, and fishery data in this application.

177 **Methods**

178 To fit multiple data types that are informative about movement (Table 1), we discuss a model
179 for instantaneous movement among discrete spatial areas representing diffusion, advection, and
180 taxis (a DAT model). We integrate this instantaneous process over discrete time-interval t to
181 define a movement matrix $\mathbf{M}(t)$ that describes the fraction of animals $m(g_2, g_1, t)$ moving from
182 cell g_1 to g_2 during interval t . We call $\mathbf{M}(t)$ the “movement matrix” and $m(g_2, g_1, t)$ a
183 “movement fraction”. We then use this movement matrix $\mathbf{M}(t)$ within a spatio-temporal model
184 that predicts numeric abundance $n(g, t)$ for each grid-cell g and time-interval (see Appendix
185 Tables A1-A4 for summary of all notation), as well as movement of individual animals and
186 associated tags among cells. Estimates of movement affect our predictions of tag locations, as
187 well as expected changes in survey abundance and fishery catch-and-effort data over time;
188 movement parameters are therefore informed by all data types and this justifies our development
189 of a joint model.

190 Grid-cells $g \in \{1, 2, \dots, n_g\}$ are evenly spaced within a spatial domain where every grid cell is
191 adjacent to one to four other grid cells (i.e., all cells are connected). We track the sequence of
192 seasons and years by index t , such that movement in each interval t can be predicted in part by

193 the season u_t or year y_t associated with interval t ; this then allows for differences in
194 environmental values and habitat preferences among seasons representing, for example,
195 spawning versus foraging preferences (e.g., Lehodey et al., 2008). Although we discretize space
196 and time in the following, future research could fit individual data sets using continuous space
197 and/or time while using the same underlying movement process and parameters, and this could
198 be appropriate for data that are informative about higher-resolution processes (e.g., electronic
199 tags).

200 **Diffusion-advection-taxis movement process**

201 We define the diffusion-advection-taxis process among discrete cells to predict an
202 instantaneous movement (a.k.a. transition) rate $m^*(g_2, g_1, t)$ from cell g_1 to cell g_2 at any
203 moment during interval t . Fish can only move between adjacent cells as time interval $\Delta t \rightarrow 0$, so
204 $m^*(g_2, g_1, t) = 0$ for any two cells that are not adjacent. We use an asterisk to indicate an
205 instantaneous rate, e.g., $\mathbf{M}^*(t)$ for the matrix of instantaneous movement rates, and subsequently
206 drop this asterisk, e.g., $\mathbf{M}(t)$, to indicate a process that is integrated over the entire time-interval.
207 Instantaneous movement matrix $\mathbf{M}^*(t)$ is a continuous-time Markov chain CTMC (e.g., Hanks et
208 al., 2015), and deriving movement matrix $\mathbf{M}(t)$ from this allows for path-dependent movement
209 between nonadjacent locations (i.e., movement between two locations is higher when there is a
210 corridor of preferred habitats between those locations).

211 We decompose instantaneous movement matrix $\mathbf{M}^*(t)$ into:

- 212 1. Diffusion matrix $\mathbf{D}^*(t)$, representing the undirected movement of animals away from their
213 present location, treated as a random-walk process;

- 214 2. Taxis matrix $\mathbf{Z}^*(t)$ representing the directed movement of animals towards preferred
215 habitats;
- 216 3. Advection matrix $\mathbf{V}^*(t)$ representing the passive transport of animals along one or more pre-
217 specified vector fields, e.g., ocean currents;

218 where:

$$m^*(g_2, g_1, t) = d^*(g_2, g_1, t) + z^*(g_2, g_1, t) + v^*(g_2, g_1, t) \quad (1)$$

219 This decomposition distinguishes taxis and advection, although other authors use these terms
220 interchangeably. Previous authors (Sibert et al., 1999; Wikle, 2003) have derived this CTMC
221 among discrete spatial areas from a partial differential equation (PDE) for movement in continuous
222 space, and we simplify our presentation by not repeating a formal derivation. Conceptually,
223 however, a partial differential equation (PDE) for diffusion specifies that density decreases where
224 the second derivative is negative (i.e., density $d(s)$ is higher at location s than the average of
225 nearby locations), and a PDE for taxis specifies that densities flow towards areas with higher
226 preference (i.e., densities decrease where preference is lower than nearby locations, and
227 correspondingly increase in those nearby preferred locations). The CTMC model replaces this
228 differential equation with analogous “movement rules” among grid cells, where diffusion
229 involves movement away from the current grid cell (i.e. diagonal elements of $\mathbf{D}^*(t)$ are
230 negative), while taxis involves movement towards grid cells with higher preference (i.e.,
231 $v^*(g_2, g_1, t)$ is positive if and only if g_2 has higher preference than g_1 in time t).

232 Previous studies for marine fishes have developed “Markov movement models” (e.g., Heifetz
233 & Fujioka, 1991; Webster et al., 2013; Hanselman et al., 2015) that estimate a pairwise

234 movement rate for each pair of locations. The diffusion-advection-taxis model differs from these
235 previous “unstructured Markov” models in two important ways:

- 236 1. Our approach uses environmental data to predict habitat preferences, thereby linking
237 movement to habitat information;
- 238 2. As a result of the first, we can define movement at fine spatial scales using a small number of
239 habitat-preference and diffusion parameters. This allows us to downscale predictions of
240 movement to much finer scales than unstructured Markov models, without an explosion in
241 the number of pairwise movement rates that must be estimated.

242 This decomposition has some similarity to that used in the Spatial Population Model SPM (Dunn
243 et al., 2015) used, for example, for Antarctic Toothfish (Mormede et al., 2014). However, SPM
244 decomposes seasonal movement probabilities into preference and/or diffusion, whereas we
245 decompose instantaneous movement rates explicitly into the three components listed. One
246 consequence of this distinction is that SPM does not account for path dependence (the impact of
247 intervening habitats on movement probabilities for two non-adjacent cells) when integrating
248 paths to calculate seasonal movement probabilities.

249 Taxis is predicted using environmental data $x(g, t, p)$ for p habitat variables, such as those
250 commonly available for species distribution (e.g., bathymetry) and climate modelling (e.g.,
251 ocean temperature). In the following, we apply basis expansion to transform these variables
252 $x(g, t, p)$ into a vector of covariates $x(g, t, k)$ used to model habitat preferences, where k
253 indexes the resulting basis functions. These transformations can involve any standard basis
254 expansion, for example, using splines, interactions, or spatial kernels, and the analyst can explore
255 alternative formulations during model development (in our software using the formula interface
256 in R).

257 Specifically, diffusion is the random movement of individuals from any cell g_1 to adjacent
 258 cells g_2 . In the following, we specify a constant diffusion rate:

$$d^*(g_2, g_1, t) = \begin{cases} e^{2\beta} & \text{if } g_1 \text{ and } g_2 \text{ are adjacent} \\ - \sum_{g' \neq g_1} d^*(g', g_1, t) & \text{if } g_1 = g_2 \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

259 where β is the log-diffusion rate, and g' (used in the 2nd row of Eq. 2) is an index used to sum
 260 across columns such that columns sum to zero. We note that future research could easily extend
 261 this notation to specify that diffusion is a function of local environmental conditions (e.g., Wikle,
 262 2003), but do not elaborate here for clarity of presentation.

263 Similarly, advection represents movement due to a vector-field, for example, the passive drift
 264 of larvae due to ocean currents and tidal flows. We envision the analyst specifying n_m vector-
 265 fields indexed by m :

$$v^*(g_2, g_1, t) = \sum_{m=1}^{n_m} \gamma(m) w(g_2, g_1, t, m), \quad (3)$$

266 where $w^*(g_2, g_1, t, m)$ is transport rate from g_1 to adjacent cell g_2 during interval t for vector-
 267 field m (e.g., oceanographic currents for passive drift), and $\gamma(m)$ are estimated parameters
 268 relating vector-field $w^*(g_2, g_1, t, m)$ to advection rate $v^*(g_2, g_1, t)$. We note that $w^*(g_2, g_1, t)$
 269 would likely be defined to conserve abundance, i.e., $\sum_{g_2=1}^{n_g} w^*(g_2, g_1, t, m) = 0$.

270 Finally, taxis represents directional movement towards preferred habitats, where habitat
 271 preference $h(g, t)$ is defined as a function of local environmental conditions:

$$h(g, t) = \sum_{k=1}^{n_k} \alpha(k) x(g, t, k), \quad (4)$$

272 where $x(g, t, k)$ is a basis-expansion of environmental layers $x(g, t, p)$, e.g., via application of
 273 splines and variable interactions, and $\alpha(k)$ are estimated environmental-preference parameters.
 274 Taxis is then defined based on local differences in the habitat-preference function:

$$z^*(g_2, g_1, t) = \begin{cases} h(g_2, t) - h(g_1, t) & \text{if } g_1 \text{ and } g_2 \text{ are adjacent} \\ - \sum_{g' \neq g_1} \{h(g_2, t) - h(g_1, t)\} & \text{if } g_1 = g_2 \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

275 We note that only the difference in preference $h(g_2, t) - h(g_1, t)$ is used, and therefore habitat-
 276 preference covariates $x(g, t, k)$ should not include an intercept term. Deriving taxis from a
 277 habitat-preference function was proposed by Brillinger (2012); it has been adapted in other
 278 studies (Preisler et al., 2013), and it appears to perform similarly to alternatives such as Langevin
 279 diffusion (Michelot et al., 2019).

280 We note that advection and taxis are both directional movement and therefore may seem
 281 similar or even redundant. However, taxis is a conservative vector field such that differences in
 282 habitat preference $h(g_1, t)$ and $h(g_2, t)$ between two locations will always be identical
 283 regardless of the path followed: this characteristic allows taxis to be uniquely defined given
 284 habitat-preference function h , such that covariates can be used flexibly to define h . By contrast,
 285 advection is not restricted to be a conservative vector field. In particular, advection can represent
 286 cyclic or chaotic movement due to passive transport in eddies, streamers, and other
 287 oceanographic features, which taxis cannot approximate due to its specification as a conservative
 288 vector field.

289 Given this definition for instantaneous movement rate $\mathbf{M}^*(t)$, we then integrate movement
 290 fraction $\mathbf{M}(t)$ over the interval starting at time t as:

$$\mathbf{M}(t) = e^{\mathbf{M}^* \Delta t}, \quad (6)$$

291 where we specifically use the matrix-exponential operator and Δt is the duration of the interval
292 starting at time t . Given that \mathbf{M}^* is a Metzler matrix (i.e., $m^*(g_2, g_1, t) \geq 0$ for all $g_2 \neq g_1$) and
293 has columns that sum to zero, $\mathbf{M}(t)$ will be non-negative (i.e., $m(g_2, g_1, t) \geq 0$) and will have
294 columns that sum to one. As a result, transformation $\mathbf{M}(t)$ conserves abundance and will have a
295 leading eigenvalue of 1.0 such that the leading eigenvector represents the hypothetical long-term
296 habitat utilization if those environmental conditions persisted indefinitely; this property of a
297 continuous-time Markov chain has been noted previously elsewhere (Wilson et al., 2018;
298 Thorson et al., 2021).

299 We illustrate these calculations in a simplified one-dimensional spatial domain to clarify
300 concepts (Fig. 1), and provide an R package on GitHub to facilitate rapid deployment for other
301 stocks (<https://github.com/James-Thorson-NOAA/ATM>) while using release 1.0.0 here. The
302 package uses a “formula” interface to allow users to specify nonlinear and interactive covariate
303 responses using covariates provided in widespread geospatial data types (Pebesma, 2019) to
304 facilitate model exploration.

305 **Spatio-temporal model for spatial distribution**

306 We next embed this DAT movement model within a spatio-temporal model for numerical
307 abundance that can be used to predict changes in distribution resulting from alternative or future
308 values of environmental layers. In this sense, our spatio-temporal DAT model represents a
309 “hybrid” species distribution model (hybrid SDM), with properties that are intermediate between
310 “correlative” spatio-temporal SDMs and “mechanistic” SDMs that include behavioral
311 information about individual movement (Dormann et al., 2012).

312 Specifically, we estimate numerical abundance $n(g, t)$ as a latent variable. This involves
 313 specifying a state-space model for $n(g, t)$ for every year t :

$$\log(\mathbf{n}(t)) = \begin{cases} \delta(t) + \boldsymbol{\varepsilon}(t) & \text{if } t = 1 \\ \delta(t) + \mathbf{M}(t-1)\mathbf{n}(t-1) + \boldsymbol{\varepsilon}(t) & \text{if } t > 1 \end{cases} \quad (7)$$

314 where $\delta(t)$ represents interannual variation in total abundance and $\boldsymbol{\varepsilon}(t)$ is a Gaussian Markov
 315 random field representing residual spatio-temporal variation in $n(g, t)$:

$$\boldsymbol{\varepsilon}(t) \sim \begin{cases} \text{MVN}(0, \sigma_0^2 \mathbf{Q}^{-1}) & \text{if } t = 1 \\ \text{MVN}(0, \sigma_\varepsilon^2 \mathbf{Q}^{-1}) & \text{if } t > 1 \end{cases} \quad (8)$$

316 where the initial process error has a different variance (σ_0^2) than the variance of subsequent
 317 process errors (σ_ε^2), and \mathbf{Q}^{-1} is an approximation to the Matérn correlation among grid cells.
 318 This Matérn correlation function represents a decline in correlation as a function of distance
 319 between any two locations \mathbf{s}_1 and \mathbf{s}_2 , and we specifically estimate a 2×2 transformation matrix
 320 \mathbf{H} such that correlations decline as a function of transformed Euclidean distance $\|\mathbf{H}(\mathbf{s}_1 - \mathbf{s}_2)\|$.
 321 This linear transformation \mathbf{H} represents geometric anisotropy and governs the rate at which
 322 correlations decline in different cardinal directions (Lindgren et al., 2011; Thorson et al., 2015).
 323 We additionally use the SPDE approximation to the correlation \mathbf{Q}^{-1} among grid cells (Lindgren
 324 et al., 2011), where inverse-correlation \mathbf{Q} in Eq. 8 is specified via sparse matrices calculated
 325 using R-INLA (Lindgren, 2012).

326 **Metrics of resulting distribution shift**

327 Many stock assessment models operate at a coarser spatial resolution than the spatial scale for
 328 environmental layers that are feasible here. To demonstrate how results from the DAT model
 329 can be adapted for use in a spatially-stratified assessment model, we therefore use “change-in-
 330 support” methods to coarsen the resolution while calculating abundance and movement among

331 strata. This process involves calculating the abundance-weighted average movement among each
 332 pair of grid-cells, while summing abundance for all grid-cells within each stratum.

333 First, we calculate abundance for each stratum o as:

$$n_{coarse}(o, t) = \sum_{g=1}^{n_g} p(g, o)n(g, t), \quad (9)$$

334 where $p(g, o)$ is the proportion of grid-cell g that is within stratum o . Next, we calculate
 335 abundance $n_{coarse}(o_2, o_1, t)$ moving from each stratum o_1 to each other stratum o_2 :

$$n_{coarse}(o_2, o_1, t) = \sum_{g_2=1}^{n_g} \sum_{g_1=1}^{n_g} n(g_1, t)p(g_2, o_2)p(g_1, o_1)m(g_2, g_1, t), \quad (10)$$

336 where coarsened movement fraction $m_{coarse}(o_2, o_1, t)$ from stratum o_1 to each other stratum o_2
 337 are then calculated by normalizing for each column:

$$m_{coarse}(o_2, o_1, t) = \frac{n(o_2, o_1, t)}{\sum_{o'=1}^{n_o} n(o', o_1, t)}, \quad (11)$$

338 where $m_{coarse}(o_2, o_1, t)$ can be used similarly to the movement rates that are calculated from
 339 previous Markov models applied to conventional tagging data (Hanselman et al., 2015; Heifetz
 340 & Fujioka, 1991).

341 We also note that either fine-scale or coarsened movement fractions ($m(g_2, g_1, t)$ and
 342 $m_{coarse}(o_2, o_1, t)$, respectively) can be used to calculate the stationary distribution for abundance
 343 $\mathbf{n}(t)$, representing the equilibrium proportion of abundance in each cell g or stratum o that
 344 would occur if current conditions persisted indefinitely. Specifically, the stationary distribution
 345 is calculated as the dominant eigenvector of $\mathbf{M}(t)$ or $\mathbf{M}_{coarse}(t)$, normalized to sum to one
 346 (noting that the eigenvalue associated with the stationary distribution is 1.0 given that movement
 347 is defined to conserve abundance). This stationary distribution can be calculated given conditions

348 in each individual year-season combination, or for the product of movement over all seasons in a
349 given year to yield the “annualized” stationary distribution.

350 **Parameter estimation**

351 We estimate parameters by identifying their values that maximize the log-likelihood of
352 available data while integrating across the probability of random effects. We approximate this
353 integral using Template Model Builder (Kristensen et al., 2016) within the R statistical
354 environment (R Core Team, 2017). We treat $\log(\mathbf{n}(t))$ as a latent process, i.e., using a state-
355 space parameterization for the Gaussian Markov random field $\boldsymbol{\epsilon}_t$ representing spatio-temporal
356 variation in Eq. 7-8 and then integrating across $\log(\mathbf{n}(t))$ by treating it as a random effect. The
357 log-likelihood is optimized using a gradient-based Nelder-Mead algorithm, followed by two
358 iterations of a Newton algorithm to decrease the final gradient. We confirm that models are
359 converged by (1) ensuring that the gradient of the log-likelihood with respect to each fixed effect
360 is <0.0001 , and (2) that the Hessian matrix is positive definite. We then calculate standard errors
361 for parameters and derived quantities using this Hessian matrix and a generalization of the delta-
362 method (Kass & Steffey, 1989), or by sampling from the joint precision matrix for fixed and
363 random effects. All parameter estimation is done using release number 1.0.0 of a new R package
364 *ATM* that is publicly available on GitHub (<https://github.com/James-Thorson-NOAA/ATM>), and
365 this paper serves as introduction for that package. *ATM* shares software dependencies (i.e.,
366 utility functions in package *FishStatsUtils*) with package *VAST* (Thorson & Barnett, 2017) to
367 facilitate comparison of parameter estimates across these two model platforms.

368 The maximum-likelihood estimation approach requires specifying a probability distribution
369 for data given parameters. We introduce this likelihood for survey data, fishery catch-per-unit-

370 effort, conventional tags, electronic tags, movement gates, and ecogeochemical tracers in
371 Appendix B.

372 **Case-study application**

373 We apply this model to three types of data for Pacific cod in the eastern and northern Bering Sea.
374 We specify a resolution involving two 6-month seasons per year (Summer: April-Sept; Winter:
375 Oct.-March) and 38 years (1982-2019), and using 100 square grid-cells each covering a
376 100 km × 100 km area to discretize the fully-connected spatial domain. The base model took
377 under two hours to estimate parameters and calculate associated standard errors on a personal
378 laptop starting from uninformative parameter values.

379 We fit three types of data to inform movement rates (see Appendix C for details):

- 380 1. Summer bottom trawl survey from 1982-2019 with fewer years sampled in the northern than
381 eastern Bering Sea;
- 382 2. Winter and summer longline fishery catch-and-effort from 1996-2019, using the total number
383 of hooks as the measure of effort a_f ;
- 384 3. Tag release and recapture locations for 2670 conventional tags from 1982-2006 and 72
385 archival tags released from 2002-2004 (Nichol et al., 2007), recovered by a mixture of
386 commercial fisheries and resource surveys. We measure fishery effort as the total number of
387 longline fishery hooks in each grid cell for a given season and year, and assume that this is
388 proportional to recapture probability conditional on the tag being present in a given cell. We
389 note that conventional and archival tags were also recovered by other fisheries and resource
390 surveys that have a different spatial footprint. We do not have access to high-resolution
391 locational information for these other fisheries, and recommend future research to reconstruct

392 this information to better interpret recapture probabilities for Pacific cod conventional tags.
393 However, an alternative hypothesis is that recapture probabilities are spatially constant, and
394 we confirm that results are qualitatively similar when a uniform distribution of fishing effort
395 is specified instead (results not shown).

396 We fit these data using two environmental layers (see Appendix Figs. D1-D3):

- 397 A. Seafloor bathymetry, previously developed for use in designating essential fish habitat (see
398 Fig. A1) and compiled from several sources (Zimmermann & Benson, 2013; Zimmermann &
399 Prescott, 2018).
- 400 B. Bottom temperature, using hindcasted values derived from the Bering10K model (Hermann
401 et al., 2016; Kearney et al., 2020). We use the simulated bottom temperature values averaged
402 across the bottom 5 m above the seafloor, and then average this across months and cells to
403 conform to the seasonal and spatial resolution of the DAT model (Appendix D Figs. D2-D3).

404 We hypothesize that Pacific cod will exhibit a nonlinear habitat preference for each
405 environmental layer. We therefore explore estimating a separate cubic basis-spline with three
406 knots (four parameters total) representing the habitat preference for each layer in each season.
407 However, preliminary exploration suggested that it was not feasible to estimate a separate
408 temperature preference for each season, so we instead used three splines (12 covariate response
409 parameters total) representing a nonlinear response to summer bathymetry, winter bathymetry,
410 and bottom temperatures.

411 *Investigating performance*

412 We investigate model performance in a variety of different ways:

- 413 1. We visualize probability integral transform (PIT) residuals (Smith, 1985) calculated from the
414 conditional simulation distribution for fishery and survey point-count data. This is useful to
415 detect departures from the specified distribution for those data.
- 416 2. We visualize the predicted and observed count of conventional tags that are detected in each
417 grid cell upon recapture, and again calculate PIT residuals from this distribution under the
418 assumption that counts follow a Poisson distribution.
- 419 3. We visualize the predicted movement of individuals from a hypothetical release location in
420 the southern middle domain, if released during the beginning of a prolonged cold stanza
421 (summer 2007) or warm stanza (summer 2014), and projecting their subsequent movement
422 given environmental conditions over the following five years.
- 423 4. We visualize predicted habitat preference $h(g, t)$ for selected warm (2002, 2017/2018) and
424 cold years (2012) during summer and winter, and also the predicted log-abundance in those
425 same seasons and years.

426 We also conduct a “self-test” simulation experiment with 25 replicates to corroborate that the
427 model can estimate parameters given the quantity of data available. In each replicate, we:

- 428 A. simulate new response values for survey and fishery catch-and-effort data as well as the
429 location of tag recaptures, conditional upon the model structure, estimated fixed effects, and
430 predicted random effects from the case-study model;
- 431 B. refit the original estimation model to these new simulated data and record the estimates of
432 habitat preference $\hat{h}(g, t)$;

- 433 C. calculate the centered habitat preference $h_{centered}(g, t) = h(g, t) - \left(\sum_{g=1}^{n_g} h(g, t)\right) / n_g$ and
 434 $\hat{h}_{centered}(g, t) = \hat{h}(g, t) - \left(\sum_{g=1}^{n_g} \hat{h}(g, t)\right) / n_g$ in recognition that only the difference in
 435 preference between two locations in a given year is used when calculating movement;
 436 D. compare estimates of centered habitat preference $h_{centered}(g, t)$ used to simulate data with
 437 estimates of habitat preference $\hat{h}_{centered}(g, t, r)$. We specifically compute the proportion of
 438 variation explained:

$$V(t) = 1 - \frac{\sum_{g=1}^{n_g} \{h_{centered}(g, t)\}^2}{\sum_{g=1}^{n_g} \{h_{centered}(g, t) - \hat{h}_{centered}(g, t)\}^2} \quad (12)$$

- 439 E. repeat steps A-D for each of 25 replicates, and compile $V(t)$ across all replicates and years.
 440 A well-performing model will result in $V(t)$ close to 1.0, while a model that cannot explain
 441 variation in preference will result in $V(t)$ close to 0.

442 The Pacific cod stock assessment has explored a spatially-stratified age- and length-
 443 structured stock assessment model implemented using Stock Synthesis (Methot & Wetzel, 2013;
 444 Thompson et al., 2020); this model separately tracks abundance-at-age in the eastern Bering Sea
 445 and northern Bering Sea (where each stratum is defined as the spatial footprint of an associated
 446 bottom trawl survey program), while estimating movement between those two strata. We
 447 therefore use the DAT model to explore three metrics for the proportion of biomass in the
 448 northern Bering Sea:

- 449 A. *Abundance*: The proportion of abundance $n_{coarse}(o, t) / \sum_{o'=1}^{n_o} n_{coarse}(o', t)$ in the eastern
 450 Bering Sea relative to total abundance (where each is defined as the spatial footprints of the
 451 eastern and northern Bering Sea bottom trawl surveys);

- 452 B. *Equilibrium*: The stationary distribution resulting from abundance-expanded movement
453 among strata, $\mathbf{M}_{coarse}(t)$, given conditions in time t ;
- 454 C. *Forward-projection*: The proportion of abundance in each stratum initialized from the
455 stationary distribution for abundance in time $t = 1$, but projected forward deterministically
456 for subsequent times using abundance-expanded movement among strata, $\mathbf{M}_{coarse}(t)$.

457 The first metric depends only upon predicted abundance $n(g, t)$ and is therefore informed
458 primarily by survey and fishery data. The second and third metrics, by contrast, are calculated
459 entirely from $\mathbf{M}_{coarse}(t)$, which in turn combines information about distribution from
460 survey/fishery data with information about diffusion and taxis. These two metrics differ in
461 whether they assume that present conditions persist indefinitely (*Equilibrium*), or using the time-
462 series of environmental conditions in each preceding time (*Forward projection*). They therefore
463 correspond to alternative definitions of prevailing environmental conditions, and are analogous
464 to “Moving average” and “Dynamic” calculations for biological reference points for use in
465 fisheries management (O’Leary, Thorson, Miller, et al., 2020).

466 Finally, we also explore the impact of including results from this model in one of the
467 assessment models (named Model 20.7) explored in the 2020 stock assessment for Bering Sea
468 Pacific cod; see Appendix E for details regarding this assessment model. To explore the impact
469 of using movement indices in this assessment, we first coarsen the spatial resolution of our
470 movement model to calculate seasonal movement $\mathbf{M}_{coarse}(t)$ between these two spatial strata
471 using Eq. 9-11. We next aggregate across both seasons, such that annual movement fractions
472 incorporate movement during summer and the following winter (i.e., summer-to-summer
473 movement). We include this coarsened and annualized movement fraction within the spatially
474 stratified Pacific cod assessment model, and extract resulting estimates of age 1+ abundance in

475 the eastern and northern Bering Sea, converting this to the fraction of total abundance in the
476 eastern Bering Sea, and compare this fraction with the survey index fraction in years with
477 consistent bottom trawl sampling across both strata (2010, 2017, 2018 and 2019). A well-
478 performing assessment model is expected to estimate a similar fraction in the eastern and
479 northern Bering Sea to what was observed in this survey.

480 *Sensitivity analyses*

481 We explore the sensitivity of model results to sequentially excluding one of the three major data
482 types: (1) conventional tags; (2) survey data; and (3) fishery catch-and-effort data. This
483 sensitivity test is designed to illustrate how each data type is influencing results. We
484 acknowledge that the model developed here differs in many ways from conventional methods for
485 analysis of tags, survey data, and fishery CPUE data. We therefore distribute the code as a
486 publicly available R package, and encourage future research comparing results to alternative
487 implementations using individual data sets and/or alternative model structures.

488 **Results**

489 By fitting to fishery catch-and-effort, survey, and tagging data (the “base model”), the diffusion-
490 advection-taxis model estimates that adult Pacific cod prefer increasing temperatures (within the
491 ranges encountered here), with a rapid decline in habitat preference for temperatures $< 0^{\circ}\text{C}$ (Fig.
492 2A). Similarly, adults show a strong preference (high peak) for depths 50-400 m during summer,
493 and a broader depth preference during winter (Fig. 2B). Diagnostics for model goodness-of-fit
494 indicate little residual spatial pattern in the match between predicted and observed recapture
495 locations (Fig. 3), and also survey and fishery residuals have quantile residuals that follow the
496 expected one-to-one line (Appendix D Fig. D4-D5). Similarly, the self-test simulation

497 experiment confirms that data are sufficient to explain over 2/3 of variation (0.69, see Fig. D6) in
498 centered habitat preferences under ideal circumstances (i.e., given that the model is correctly
499 specified).

500 The estimated habitat preference function (and resulting taxis) shows only small differences
501 in movement for individuals during a cold stanza (2007-2012) and warm stanza (2014-2019)
502 (Fig. 4, comparing columns), including between summer and the subsequent winter distribution
503 (Fig. 4, comparing 1st and 2nd rows). Despite this broad similarity in movement between warm
504 and cold stanzas, the model does predict slight increases in northward movement during warm
505 stanzas, e.g., elevated movement probabilities from the southern middle domain release location
506 to south of St. Lawrence Island for a release in 2014 relative to that in 2007 (comparing bottom
507 row of Fig. 4). Differences in habitat preference are more clearly illustrated by comparing the
508 estimated winter preference function in warm years (2002/2019) and a cold year (2012) (Fig. 5,
509 2nd column). These differences are above the approximately 0.1 standard error in estimated
510 preferences (see Appendix D Fig. D7). The predicted densities show greater differences among
511 years, where the northern Bering Sea has elevated densities in 2017/2019 relative to 2002/2012
512 in both summer and winter seasons (Fig. 5, 3rd and 4th columns).

513 Next, we visualize movement between the eastern and northern Bering Sea to understand
514 whether we can predict the recent increase in proportion in the northern stratum of the most
515 recent stock assessment. The model predicts that 30-70% of individuals in the northern Bering
516 Sea would move back to the eastern Bering Sea during a given year from 1982-2012 (Fig. 6).
517 This proportion started dropping in 2013, and was 5-10% in 2017-2019, with an associated but
518 smaller increase in the proportion predicted to move from the eastern to northern Bering Sea
519 strata. Given those movement fractions, the *Equilibrium* metric for proportion closely matched

520 the *Abundance* metric for the proportion in the northern stratum, reaching 20-40% from 2017-
521 2019 (see Methods for detailed calculation). As expected, the *Forward-projection* metric lagged
522 behind the *Equilibrium* metric, and reached 20% in the northern stratum by 2019. We therefore
523 conclude that including abundance-expanded movement among-strata within a spatially-
524 stratified assessment model would capture some (but not all) of the northward shift that is
525 apparent from survey and fishery data.

526 As a sensitivity analysis, we examine these same results for a model that sequentially drops
527 each major data type (fishery catch-and-effort, survey, or tagging data). This shows two major
528 points. First, removing tagging data results in a much smaller estimate of diffusion rate (Fig. 7,
529 bottom-left panel). This then results in lower movement among strata, in turn causing a slower
530 response of the *Forward-projection* metric to changing environmental conditions and a degraded
531 match between that metric and the *Abundance* metric (Fig. 7, bottom-right panel). Second,
532 removing survey data results in an implausible increase in the estimated density in the northern
533 Bering Sea (Fig. 7, 2nd row right column). By contrast, removing fishery data results in
534 relatively little change in the qualitative results (Fig. 7, top row), suggesting that fishery catch-
535 and-effort data are less important than survey and tagging data for reconstructing movement for
536 this population.

537 Finally, we demonstrate the impact of including estimates of annual movement fractions in
538 the most recent spatially stratified stock assessment model for Pacific cod (Fig. 8). The existing
539 “two-box” model (Fig. 8 grey line) estimates a high proportion (10-50%) of abundance in the
540 northern Bering Sea throughout the entire period, despite a high-quality systematic survey in
541 2010 detecting almost no Pacific cod in that area and earlier surveys in the 1980s similarly
542 recording low densities in the NBS (Fig. 8 black bullets). This existing assessment model

543 attributes the disagreement between predicted abundance and survey data to low catchability for
544 cod in the northern Bering Sea prior to 2017. By contrast, the same model using movement
545 fractions estimated here predicts that almost all abundance is in the eastern Bering Sea
546 throughout 1990-2010 (Fig. 8 blue line). It then predicts an increasing proportion after 2013
547 with patterns that closely match the “forward-projected” proportions shown previously (Fig. 6).
548 We therefore conclude (1) that movement fractions are easily included in the widely used stock
549 assessment platform used in this assessment, and (2) that this information can greatly impact
550 predictions of population movement for this commercially important stock. However, the age-
551 structured model using movement rates estimated here still substantially underestimates the
552 proportion observed in the northern Bering Sea in 2017-2019 based on survey data (Fig. 8 black
553 bullets).

554 **Discussion**

555 We have demonstrated a “hybrid” species distribution modelling framework (sensu Dormann
556 et al., 2012) that has characteristics in common with both correlative SDMs (i.e., fitting to data
557 statistically using covariates) and mechanistic methods (i.e., including individual movement
558 processes). We used this hybrid-SDM to integrate climate and habitat variables with data from
559 conventional tags, fishery, and survey operations to estimate fine-scale seasonal movement
560 patterns for a marine species, and also presented future extensions to integrate movement gates,
561 archival tags, ecogeochemical tracers, and other emerging data types. We also demonstrated
562 how to coarsen the spatial and seasonal resolution to calculate annualized movement fractions
563 among large spatial strata that can be easily integrated within existing spatially stratified stock
564 assessment models. Finally, we demonstrated the benefits of this approach for Bering Sea Pacific
565 cod, a commercially important fish stock experiencing climate-driven northward range shifts. In

566 the following, we next (1) discuss avenues to improve our ecological understanding of Pacific
567 cod in particular, (2) summarize improvements of this instantaneous diffusion-advection-taxis
568 approach relative to previous methods integrating tag and survey data, (3) review how the DAT
569 model can be connected to IPM and stock assessment models, and (4) envision how this
570 approach could be used to optimize ongoing deployment of tagging efforts.

571 We have used Pacific cod in the Bering Sea to demonstrate our integrated approach, but note
572 that we have restricted environmental preferences to bathymetry and seasonal bottom
573 temperature. We recommend that future research explore including additional covariates
574 influencing seasonal and interannual variability in Pacific cod distribution. The processes that
575 drive seasonal and interannual variation in distribution likely differ between feeding and
576 spawning seasons, and the timing of spawning migration likely vary between years. Variation in
577 the winter spawning distribution is influenced by the timing and extent of ice coverage and
578 bottom temperatures (Neidetcher et al., 2014), while variation in spatial distribution outside of
579 the spawning season is likely influenced by the variability in distribution of forage. The
580 distribution of forage fishes, in turn, is associated with bottom depth and temperature, but also
581 salinity, water column stability, concentration of chlorophyll-a, zooplankton biomass and
582 composition, as well as annual indices of regional oceanography (Parker-Stetter et al., 2016).
583 Future developments of the DAT model could incorporate these regional oceanographic indices
584 using a spatially varying response when estimating habitat preferences (Thorson, 2019b). The
585 timing of Pacific cod spawning is also variable (Neidetcher et al., 2014), perhaps in response to
586 the narrow window of temperatures resulting in optimal egg hatch success (Laurel & Rogers,
587 2020), and this could be resolved using finer seasonal increments (e.g., Thorson et al., 2020).

588 The DAT model builds upon existing advection-diffusion-reaction (ADR) analyses used in
589 fisheries. For example, SEAPODYM estimates habitat preference separately for spawning and
590 feeding stages based on tagging data, which are then used to drive advection within an advective-
591 diffusive movement model (Lehodey et al., 2008; Senina et al., 2019). Similarly, SPM integrates
592 tagging and survey data to fit an age-structured population model, while projecting movement
593 based on parameters representing movement probabilities and an estimated seasonal movement
594 kernel (Dunn et al., 2015). However, we are not aware of other fisheries models that estimated
595 advection-diffusion movement while also incorporating spatially correlated process errors and
596 fitting to point-count data from fisheries and surveys. Identifying environmental associations
597 using a hybrid SDM can be used directly within fisheries management, to identify long-term
598 “essential fish habitat” and short-term bycatch hotspots (Maxwell et al., 2015; Rooper et al.,
599 2020). Parameterizing hybrid SDMs is increasingly feasible given recent advances in biologging,
600 as well as detailed habitat and oceanographic mapping efforts (Lowerre-Barbieri et al., 2019).

601 As demonstrated here, fine-scale predictions of movement can be coarsened to estimate
602 seasonal or annualized movement fractions among larger spatial strata, and these movement
603 fractions can be inputted as data into spatial IPMs or used to specify priors on movement
604 parameters. These movement fractions are conditioned upon both (1) the estimated advective-
605 diffusion movement process and (2) the fine-scale distribution of numerical abundance.
606 Estimating fine-scale distribution is facilitated by fitting survey and fishery data, and spatially
607 correlated process errors are useful to accurately incorporate interannual variation in distribution
608 (Thorson, 2019a). In the future, we envision that the DAT model itself could be merged into an
609 integrated stock assessment model to provide improved estimates of movement among coarsened
610 strata, while enabling the incorporation of a wider array of data sources. Along these lines, we

611 recommend future research to extend the DAT model to better account for stage based
612 movement dynamics, such as migratory behavior (i.e., spawning migrations); these could be
613 modeled by estimating separate preference functions for different stages. We also recommend
614 ongoing research to incorporate spatial variation in demographics (e.g., habitat-specific fishing
615 mortality rates) within the DAT model.

616 Once a hybrid SDM has been developed and validated, analysts can then use the
617 parameterized model as a simulation tool to optimize the location and timing of tag releases to
618 maximize information gain with respect to an explicit management goal. Conventional tags are
619 often deployed systematically with large sample sizes as a component of a standardized fisheries
620 survey, randomly tagging fish proportionally to catch through space and time (e.g., Echave et al.,
621 2013). This method can be useful for understanding general movement patterns when
622 developing an initial hybrid SDM, but recapture rates are typically low (Shimada & Kimura,
623 1994) resulting in little statistical power for estimating processes affecting movement within
624 small subareas or habitats of particular management interest. By contrast, newer archival and
625 satellite tags remain expensive, and typically are deployed in targeted campaigns without a
626 population-scale probabilistic design. We recommend using a hybrid SDM as an operating
627 model within a formal optimization exercise, using results (along with other operational and
628 management constraints) to inform the deployment of targeted tagging efforts using additional
629 electronic tags. We hypothesize that times and locations with historically less data or high
630 variability in a fitted hybrid SDM could be a useful guidepost for future tagging efforts, but
631 formal optimization may also yield surprises contrary to this pattern. For example, sampling
632 optimization using occupancy models has recommended increasing sampling in locations with

633 intermediate densities (rather than proportional to estimation variance) to better determine range
634 limits (e.g., Reich et al., 2018).

635 In conclusion, we foresee a tremendous decrease in cost and increase in deployment for
636 technologies tracking movement during the “Biologging decade.” We also see a growing need
637 to integrate these novel movement-sampling technologies and programs with conventional data
638 types. This synthesis is needed, for example, to provide “best available science” advice for
639 fisheries management and marine spatial planning, including the overlap between the changing
640 footprint of human activities and changing seasonal and spatial habitat utilization. We therefore
641 encourage researchers to develop a “community of practice” using diffusion-advection-taxis
642 models to link individual movement to habitat and climate variables within hybrid SDMs.
643 Defining annual movement from instantaneous habitat preference and diffusion parameters is
644 necessary to integrate data with large differences in temporal frequencies, e.g., daily records
645 (archival tags, movement gates, and selection experiments) and seasonal records (surveys and
646 conventional tags). A common model then allows for improved sharing of information among
647 sampling programs as well as among species (based on similarities in traits and phylogeny), and
648 promises to allow improved learning between marine regions and research communities
649 worldwide (Thorson et al., 2021).

650 **Data Availability Statement**

651 Survey catches, tagging data, and environmental layers (depth and seasonal bottom
652 temperatures) are publicly available from the Alaska Fisheries Science Center, and are repositied
653 as an example in R package ATM (<https://github.com/James-Thorson-NOAA/ATM>). Fishery
654 catch-per-unit-effort (CPUE) data are private information, and are available from the Alaska
655 Fisheries Science Center upon request and suitable data-sharing agreement. The aggregated

656 fishery CPUE data used here are repositied in R package ATM, after excluding grid cells and
657 intervals with three or fewer samples. The fishery effort data are private information, and again
658 are available from the Alaska Fisheries Science Center upon request and suitable data-sharing
659 agreement. See Appendix C for detailed description of the data used.

660 **Acknowledgments**

661 We thank many scientists for helpful discussions regarding this research, including L. Britt, S.
662 McDermott, P. Conn, and D. Johnson. JT is grateful to K. Kristensen and H. Skaug for their
663 prior guidance regarding diffusion-taxis models, and the developers of Template Model Builder
664 for making these ambitious models possible. We thank many others for contributing to the data
665 used here, including: S. McDermott, D. Bryan, D. Nichol, and others for the archival tagging
666 efforts; M. Zimmermann, J. Pirtle, C. Rooper, and others for the bathymetric layer; countless
667 fishery observers and portside samplers for the longline fishery CPUE data; A. Hermann, W.
668 Cheng, and D. Pilcher for the ROMS bottom temperature layer; and B. Lauth, S. Kotwicki and
669 others for the bottom trawl survey data. We thank C. Braun, P. Conn, E. Hazen, B. Gardner, M,
670 Mangel, and one anonymous reviewer for helpful comments on an earlier draft. This publication
671 is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO)
672 under NOAA Cooperative Agreement NA15OAR4320063, Contribution No. 2020-1135.

673 **Literature**

- 674 Berger, A. M., Goethel, D. R., Lynch, P. D., Quinn, T., Mormede, S., McKenzie, J., & Dunn, A. (2017). Space
675 oddity: The mission for spatial integration. *Canadian Journal of Fisheries and Aquatic Sciences*,
676 74(11), 1698–1716. <https://doi.org/10.1139/cjfas-2017-0150>
- 677 Brillinger, D. R. (2012). Learning a Potential Function From a Trajectory. In *Selected Works of David*
678 *Brillinger* (pp. 361–364). Springer, New York, NY. [https://doi.org/10.1007/978-1-4614-1344-](https://doi.org/10.1007/978-1-4614-1344-8_20)
679 [8_20](https://doi.org/10.1007/978-1-4614-1344-8_20)
- 680 Bruneel, S., Gobeyn, S., Verhelst, P., Reubens, J., Moens, T., & Goethals, P. (2018). Implications of
681 movement for species distribution models - Rethinking environmental data tools. *Science of The*
682 *Total Environment*, 628–629, 893–905. <https://doi.org/10.1016/j.scitotenv.2018.02.026>
- 683 Cao, J., Thorson, J. T., Punt, A. E., & Szuwalski, C. (2020). A novel spatiotemporal stock assessment
684 framework to better address fine-scale species distributions: Development and simulation
685 testing. *Fish and Fisheries*, 21(2), 350–367. <https://doi.org/10.1111/faf.12433>
- 686 Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X.,
687 Römermann, C., Schröder, B., & Singer, A. (2012). Correlation and process in species distribution
688 models: bridging a dichotomy. *Journal of Biogeography*, 39(12), 2119–2131.
689 <https://doi.org/10.1111/j.1365-2699.2011.02659.x>
- 690 Dunn, A., Rasumussen, S., & Mormede, S. (2015). *Spatial Population Model User Manual* (NIWA
691 Technical Report 138; p. 208). National Institute of Water & Atmospheric Research Ltd.
- 692 Echave, K., Hanselman, D. H., & Maloney, N. E. (2013). *Report to industry on the Alaska sablefish tag*
693 *program, 1972-2012* (NOAA Tech. Memo. NMFS-AFSC-254; p. 47).
- 694 Fournier, D. A., Hampton, J., & Sibert, J. R. (1998). MULTIFAN-CL: a length-based, age-structured model
695 for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*.

696 *Canadian Journal of Fisheries and Aquatic Sciences*, 55(9), 2105–2116.
697 <https://doi.org/10.1139/f98-100>

698 Fox, J., & Weisberg, S. (2018). *An R companion to applied regression* (2nd ed.). Sage Publications.

699 Goethel, D. R., Bosley, K. M., Langseth, B. J., Deroba, J. J., Berger, A. M., Hanselman, D. H., & Schueller, A.
700 M. (2021). Where do you think you're going? Accounting for ontogenetic and climate-induced
701 movement in spatially stratified integrated population assessment models. *Fish and Fisheries*,
702 22(1), 141–160. <https://doi.org/10.1111/faf.12510>

703 Hampton, J., & Fournier, D. A. (2001). A spatially disaggregated, length-based, age-structured population
704 model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. *Marine*
705 *and Freshwater Research*, 52(7), 937–963. <https://doi.org/10.1071/mf01049>

706 Hanks, E. M., Hooten, M. B., & Alldredge, M. W. (2015). Continuous-time discrete-space models for
707 animal movement. *The Annals of Applied Statistics*, 9(1), 145–165. [https://doi.org/10.1214/14-](https://doi.org/10.1214/14-AOAS803)
708 AOAS803

709 Hanselman, D. H., Heifetz, J., Echave, K. B., & Dressel, S. C. (2015). Move it or lose it: movement and
710 mortality of sablefish tagged in Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(2),
711 238–251. <https://doi.org/10.1139/cjfas-2014-0251>

712 Hauser, D. D. W., Laidre, K. L., & Stern, H. L. (2018). Vulnerability of Arctic marine mammals to vessel
713 traffic in the increasingly ice-free Northwest Passage and Northern Sea Route. *Proceedings of*
714 *the National Academy of Sciences*, 115(29), 7617–7622.
715 <https://doi.org/10.1073/pnas.1803543115>

716 Hays, G. C., Bailey, H., Bograd, S. J., Bowen, W. D., Campagna, C., Carmichael, R. H., Casale, P., Chiaradia,
717 A., Costa, D. P., Cuevas, E., Nico de Bruyn, P. J., Dias, M. P., Duarte, C. M., Dunn, D. C., Dutton, P.
718 H., Esteban, N., Friedlaender, A., Goetz, K. T., Godley, B. J., ... Sequeira, A. M. M. (2019).

719 Translating Marine Animal Tracking Data into Conservation Policy and Management. *Trends in*
720 *Ecology & Evolution*, 34(5), 459–473. <https://doi.org/10.1016/j.tree.2019.01.009>

721 Heifetz, J., & Fujioka, J. T. (1991). Movement dynamics of tagged sablefish in the northeastern Pacific.
722 *Fisheries Research*, 11(3), 355–374. [https://doi.org/10.1016/0165-7836\(91\)90009-5](https://doi.org/10.1016/0165-7836(91)90009-5)

723 Hermann, A. J., Gibson, G. A., Bond, N. A., Curchitser, E. N., Hedstrom, K., Cheng, W., Wang, M., Cokelet,
724 E. D., Stabeno, P. J., & Aydin, K. (2016). Projected future biophysical states of the Bering Sea.
725 *Deep-Sea Research Part II: Topical Studies in Oceanography*, 134, 30–47.
726 <https://doi.org/10.1016/j.dsr2.2015.11.001>

727 Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-time correlated random
728 walk model for animal telemetry data. *Ecology*, 89(5), 1208–1215. [https://doi.org/10.1890/07-](https://doi.org/10.1890/07-1032.1)
729 1032.1

730 Kass, R. E., & Steffey, D. (1989). Approximate Bayesian inference in conditionally independent
731 hierarchical models (parametric empirical bayes models). *Journal of the American Statistical*
732 *Association*, 84(407), 717–726. <https://doi.org/10.2307/2289653>

733 Kearney, K., Hermann, A., Cheng, W., Ortiz, I., & Aydin, K. (2020). A coupled pelagic–benthic–sympagic
734 biogeochemical model for the Bering Sea: documentation and validation of the BESTNPZ model
735 (v2019.08.23) within a high-resolution regional ocean model. *Geoscientific Model Development*,
736 13(2), 597–650. <https://doi.org/10.5194/gmd-13-597-2020>

737 Kéry, M., & Schaub, M. (2021). *Integrated Population Models: Theory and Ecological Applications with R*
738 *and JAGS* (1st edition). Academic Press.

739 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation
740 and Laplace approximation. *Journal of Statistical Software*, 70(5), 1–21.
741 <https://doi.org/10.18637/jss.v070.i05>

742 Kristensen, K., Thygesen, U. H., Andersen, K. H., & Beyer, J. E. (2014). Estimating spatio-temporal
743 dynamics of size-structured populations. *Canadian Journal of Fisheries and Aquatic Sciences*,
744 71(2), 326–336. <https://doi.org/10.1139/cjfas-2013-0151>

745 Lam, C. H., Nielsen, A., & Sibert, J. R. (2008). Improving light and temperature based geolocation by
746 unscented Kalman filtering. *Fisheries Research*, 91(1), 15–25.
747 <https://doi.org/10.1016/j.fishres.2007.11.002>

748 Laurel, B. J., & Rogers, L. A. (2020). Loss of spawning habitat and prerecruits of Pacific cod during a Gulf
749 of Alaska heatwave. *Canadian Journal of Fisheries and Aquatic Sciences*.
750 <https://doi.org/10.1139/cjfas-2019-0238>

751 Laurel, B. J., Stoner, A. W., & Hurst, T. P. (2007). Density-dependent habitat selection in marine flatfish:
752 the dynamic role of ontogeny and temperature. *Marine Ecology Progress Series*, 338, 183–192.

753 Lee, M. S. Y., & Palci, A. (2015). Morphological Phylogenetics in the Genomic Age. *Current Biology*,
754 25(19), R922–R929. <https://doi.org/10.1016/j.cub.2015.07.009>

755 Lehodey, P., Senina, I., & Murtugudde, R. (2008). A spatial ecosystem and populations dynamics model
756 (SEAPODYM) – Modeling of tuna and tuna-like populations. *Progress in Oceanography*, 78(4),
757 304–318. <https://doi.org/10.1016/j.pocean.2008.06.004>

758 Lindgren. (2012). Continuous domain spatial models in R-INLA. *The ISBA Bulletin*, 19(4), 14–20.

759 Lindgren, Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov
760 random fields: the stochastic partial differential equation approach. *Journal of the Royal*
761 *Statistical Society: Series B (Statistical Methodology)*, 73(4), 423–498.
762 <https://doi.org/10.1111/j.1467-9868.2011.00777.x>

763 Lowerre-Barbieri, S. K., Kays, R., Thorson, J. T., & Wikelski, M. (2019). The ocean’s movescape: fisheries
764 management in the bio-logging decade (2018–2028). *ICES Journal of Marine Science*, 76(2), 477–
765 488. <https://doi.org/10.1093/icesjms/fsy211>

766 Maunder, M. N., & Punt, A. E. (2013). A review of integrated analysis in fisheries stock assessment.
767 *Fisheries Research*, 142, 61–74. <https://doi.org/10.1016/j.fishres.2012.07.025>

768 Maureaud, A. A., Frelat, R., Pécuchet, L., Shackell, N., Mérigot, B., Pinsky, M. L., Amador, K., Anderson, S.
769 C., Arkhipkin, A., Auber, A., Barri, I., Bell, R. J., Belmaker, J., Beukhof, E., Camara, M. L., Guevara-
770 Carrasco, R., Choi, J., Christensen, H. T., Conner, J., ... Thorson, J. T. (2021). Are we ready to track
771 climate-driven shifts in marine species across international boundaries? - A global survey of
772 scientific bottom trawl data. *Global Change Biology*, 27(2), 220–236.
773 <https://doi.org/10.1111/gcb.15404>

774 Maxwell, S. M., Hazen, E. L., Lewison, R. L., Dunn, D. C., Bailey, H., Bograd, S. J., Briscoe, D. K., Fossette,
775 S., Hobday, A. J., Bennett, M., Benson, S., Caldwell, M. R., Costa, D. P., Dewar, H., Eguchi, T.,
776 Hazen, L., Kohin, S., Sippel, T., & Crowder, L. B. (2015). Dynamic ocean management: Defining
777 and conceptualizing real-time management of the ocean. *Marine Policy*, 58, 42–50.
778 <https://doi.org/10.1016/j.marpol.2015.03.014>

779 Methot, R. D., & Wetzel, C. R. (2013). Stock synthesis: A biological and statistical framework for fish
780 stock assessment and fishery management. *Fisheries Research*, 142, 86–99.

781 Michelot, T., Gloaguen, P., Blackwell, P. G., & Étienne, M.-P. (2019). The Langevin diffusion as a
782 continuous-time model of animal movement and habitat selection. *Methods in Ecology and*
783 *Evolution*, 10(11), 1894–1907. <https://doi.org/10.1111/2041-210X.13275>

784 Mormede, S., Dunn, A., Hanchet, S. M., & Parker, S. (2014). Spatially explicit population dynamics
785 models for Antarctic toothfish in the Ross Sea region. *CCAMLR Science*, 21, 19–37.

786 Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A
787 movement ecology paradigm for unifying organismal movement research. *Proceedings of the*
788 *National Academy of Sciences*, 105(49), 19052–19059.
789 <https://doi.org/10.1073/pnas.0800375105>

790 Neidetcher, S. K., Hurst, T. P., Ciannelli, L., & Logerwell, E. A. (2014). Spawning phenology and geography
791 of Aleutian Islands and eastern Bering Sea Pacific cod (*Gadus macrocephalus*). *Deep Sea*
792 *Research Part II: Topical Studies in Oceanography*, 109, 204–214.
793 <https://doi.org/10.1016/j.dsr2.2013.12.006>

794 Nichol, D. G., Honkalehto, T., & Thompson, G. G. (2007). Proximity of Pacific cod to the sea floor: Using
795 archival tags to estimate fish availability to research bottom trawls. *Fisheries Research*, 86(2),
796 129–135. <https://doi.org/10.1016/j.fishres.2007.05.009>

797 O’Leary, C. A., Thorson, J. T., Ianelli, J. N., & Kotwicki, S. (2020). Adapting to climate-driven distribution
798 shifts using model-based indices and age composition from multiple surveys in the walleye
799 pollock (*Gadus chalcogrammus*) stock assessment. *Fisheries Oceanography*, 29(6), 541–557.
800 <https://doi.org/10.1111/fog.12494>

801 O’Leary, C. A., Thorson, J. T., Miller, T. J., & Nye, J. A. (2020). Comparison of multiple approaches to
802 calculate time-varying biological reference points in climate-linked population-dynamics models.
803 *ICES Journal of Marine Science*, 77(3), 930–941. <https://doi.org/10.1093/icesjms/fsz215>

804 Parker-Stetter, S., Urmy, S., Horne, J., Eisner, L., & Farley, E. (2016). Factors affecting summer
805 distributions of Bering Sea forage fish species: Assessing competing hypotheses. *Deep Sea*
806 *Research Part II: Topical Studies in Oceanography*, 134, 255–269.
807 <https://doi.org/10.1016/j.dsr2.2016.06.013>

808 Pebesma, E. (2019). stars: Spatiotemporal Arrays, Raster and Vector Data Cubes. *R Package Version 0.4-*
809 *0*.

810 Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., & Levin, S. A. (2013). Marine taxa track local
811 climate velocities. *Science*, 341(6151), 1239–1242.

812 Preisler, H. K., Ager, A. A., & Wisdom, M. J. (2013). Analyzing animal movement patterns using potential
813 functions. *Ecosphere*, 4(3), art32. <https://doi.org/10.1890/ES12-00286.1>

814 R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for
815 Statistical Computing. <https://www.R-project.org/>

816 Reich, B. J., Pacifici, K., & Stallings, J. W. (2018). Integrating auxiliary data in optimal spatial design for
817 species distribution modelling. *Methods in Ecology and Evolution*, 9(6), 1626–1637.
818 <https://doi.org/10.1111/2041-210X.13002>

819 Rooper, C. N., Ortiz, I., Hermann, A. J., Laman, N., Cheng, W., Kearney, K., & Aydin, K. (2020). Predicted
820 shifts of groundfish distribution in the Eastern Bering Sea under climate change, with
821 implications for fish populations and fisheries management. *ICES Journal of Marine Science*,
822 *fsaa215*. <https://doi.org/10.1093/icesjms/fsaa215>

823 Santora, J. A., Mantua, N. J., Schroeder, I. D., Field, J. C., Hazen, E. L., Bograd, S. J., Sydeman, W. J., Wells,
824 B. K., Calambokidis, J., Saez, L., Lawson, D., & Forney, K. A. (2020). Habitat compression and
825 ecosystem shifts as potential links between marine heatwave and record whale entanglements.
826 *Nature Communications*, 11(1), 536. <https://doi.org/10.1038/s41467-019-14215-w>

827 Senina, I., Lehodey, P., Sibert, J., & Hampton, J. (2019). Integrating tagging and fisheries data into a
828 spatial population dynamics model to improve its predictive skills. *Canadian Journal of Fisheries*
829 *and Aquatic Sciences*, 77(3), 576–593. <https://doi.org/10.1139/cjfas-2018-0470>

830 Shertzer, K. W., & Bacheler, N. M. (2020). Estimating population abundance at a site in the open ocean:
831 combining information from conventional and telemetry tags with application to gray triggerfish
832 (*Balistes capriscus*) - Canadian Journal of Fisheries and Aquatic Sciences. *Canadian Journal of*
833 *Fisheries and Aquatic Sciences*, 77(1), 34–43.

834 Shimada, A. M., & Kimura, D. K. (1994). Seasonal movements of Pacific cod, *Gadus macrocephalus*, in
835 the eastern Bering Sea and adjacent waters based on tag-recapture data. *Fish Bull*, 92, 800–816.

836 Sibert, J. R., Hampton, J., Fournier, D. A., & Bills, P. J. (1999). An advection–diffusion–reaction model for
837 the estimation of fish movement parameters from tagging data, with application to skipjack

838 tuna (*Katsuwonus pelamis*). *Canadian Journal of Fisheries and Aquatic Sciences*, 56(6), 925–938.
839 <https://doi.org/10.1139/f99-017>

840 Smith, J. Q. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting*, 4(3),
841 283–291. <https://doi.org/10.1002/for.3980040305>

842 Spies, I., Gruenthal, K. M., Drinan, D. P., Hollowed, A. B., Stevenson, D. E., Tarpey, C. M., & Hauser, L.
843 (2020). Genetic evidence of a northward range expansion in the eastern Bering Sea stock of
844 Pacific cod. *Evolutionary Applications*, 13(2), 362–375. <https://doi.org/10.1111/eva.12874>

845 Thompson, G., Conner, J., Shotwell, K., Fissel, B., Hurst, T., Laurel, B., Rogers, L., & Siddon, E. (2020).
846 *Assessment of the Pacific cod stock in the Eastern Bering Sea. In Stock assessment and fishery*
847 *evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands [NPFMC*
848 *Bering Sea and Aleutian Islands SAFE]. North Pacific Fishery Management Council.*
849 <https://archive.afsc.noaa.gov/refm/docs/2019/EBSpcod.pdf>

850 Thorson, J. T. (2019a). Forecast skill for predicting distribution shifts: A retrospective experiment for
851 marine fishes in the Eastern Bering Sea. *Fish and Fisheries*, 20(1), 159–173.
852 <https://doi.org/10.1111/faf.12330>

853 Thorson, J. T. (2019b). Measuring the impact of oceanographic indices on species distribution shifts: The
854 spatially varying effect of cold-pool extent in the eastern Bering Sea. *Limnology and*
855 *Oceanography*, 64(6), 2632–2645. <https://doi.org/10.1002/lno.11238>

856 Thorson, J. T. (2020). Predicting recruitment density dependence and intrinsic growth rate for all fishes
857 worldwide using a data-integrated life-history model. *Fish and Fisheries*, 21(2), 237–251.
858 <https://doi.org/10.1111/faf.12427>

859 Thorson, J. T., Adams, C. F., Brooks, E. N., Eisner, L. B., Kimmel, D. G., Legault, C. M., Rogers, L. A., &
860 Yasumiishi, E. M. (2020). Seasonal and interannual variation in spatio-temporal models for index

861 standardization and phenology studies. *ICES Journal of Marine Science*, 77(5), 1879–1892.
862 <https://doi.org/10.1093/icesjms/fsaa074>

863 Thorson, J. T., & Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution
864 shifts using single- and multispecies models of fishes and biogenic habitat. *ICES Journal of*
865 *Marine Science*, 74(5), 1311–1321. <https://doi.org/10.1093/icesjms/fsw193>

866 Thorson, J. T., Hermann, A. J., Siwicke, K., & Zimmermann, M. (2021). Grand challenge for habitat
867 science: stage-structured responses, nonlocal drivers, and mechanistic associations among
868 habitat variables affecting fishery productivity. *ICES Journal of Marine Science*, fsaa236.
869 <https://doi.org/10.1093/icesjms/fsaa236>

870 Thorson, J. T., Shelton, A. O., Ward, E. J., & Skaug, H. J. (2015). Geostatistical delta-generalized linear
871 mixed models improve precision for estimated abundance indices for West Coast groundfishes.
872 *ICES Journal of Marine Science: Journal Du Conseil*, 72(5), 1297–1310.
873 <https://doi.org/10.1093/icesjms/fsu243>

874 Webster, R. A., Clark, W. G., Leaman, B. M., & Forsberg, J. E. (2013). Pacific halibut on the move: a
875 renewed understanding of adult migration from a coastwide tagging study. *Canadian Journal of*
876 *Fisheries and Aquatic Sciences*. <https://doi.org/10.1139/cjfas-2012-0371>

877 Wikle, C. K. (2003). Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes.
878 *Ecology*, 84(6), 1382–1394. [https://doi.org/10.1890/0012-](https://doi.org/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2)
879 [9658\(2003\)084\[1382:HBMFPT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2)

880 Wilson, K., Hanks, E., & Johnson, D. (2018). Estimating animal utilization densities using continuous-time
881 Markov chain models. *Methods in Ecology and Evolution*, 9(5), 1232–1240.
882 <https://doi.org/10.1111/2041-210X.12967>

883 Zimmermann, M., & Benson, J. L. (2013). *Smooth sheets: How to work with them in a GIS to derive*
884 *bathymetry, features and substrates* (NOAA Technical Memorandum NMFS-AFSC-249; p. 52). US

885 Department of Commerce, National Oceanic and Atmospheric Administration, National Marine
886 Fisheries Service, Alaska Fisheries Science Center.

887 Zimmermann, M., & Prescott, M. M. (2018). Bathymetry and canyons of the eastern Bering Sea slope.
888 *Geosciences*, 8(5), 184. <https://doi.org/10.3390/geosciences8050184>

889 Zipkin, E. F., Rossman, S., Yackulic, C. B., Wiens, J. D., Thorson, J. T., Davis, R. J., & Grant, E. H. C. (2017).
890 Integrating count and detection–nondetection data to model population dynamics. *Ecology*,
891 *98*(6), 1640–1650. <https://doi.org/10.1002/ecy.1831>

892

893

894 Table 1: List of some (but not all) categories of information available to estimate movement, illustrating the potential benefit of a
 895 data-integrated movement model that can assimilate a wide variety of data types.

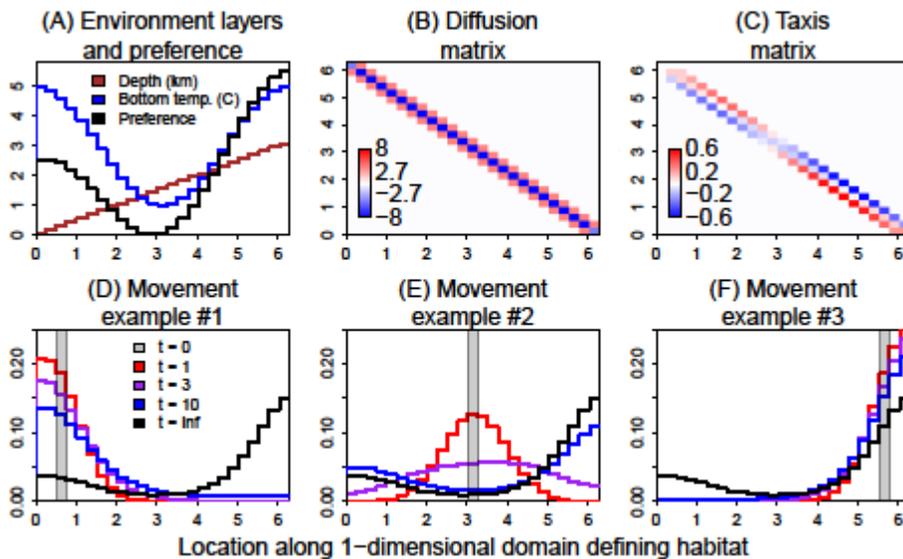
Data type	Examples	Benefits	Difficulties
Conventional tags	Floy tags for sablefish	Widely used and low cost per tag	Low return rates; Dependent upon distribution of fishing effort
Survey data	Bottom trawl surveys	Available in many regions worldwide (Maureaud et al., 2021)	Reveals movement indirectly via habitat utilization
Fishery data	Fishery catch-per-unit-effort; Fishery effort	Available in regions without surveys	Reveals movement indirectly via habitat utilization Interpretation depends upon correct assumptions
Archival tags	mPAT; etc.	Provide high-resolution information about environmental conditions in utilized habitat	Relatively high cost per tag Expensive to process output
Movement gates	Upward facing acoustics Weirs	High temporal-frequency Directly measures flux (aggregate movement)	Few technologies for use in marine environment Small spatial coverage
Selection experiments	Laboratory selection experiments	Robust understanding of causal relationship between environment and movement	Difficult to “scale-up” results from laboratory to basin scales
Chemical markers	Stable isotope ratios compared with isoscapes; trace element suites	Retrospective study; useful for untaggable fish; cheaper than tags; elucidates behavioral differences	Non-lethal methods reduce temporal resolution (muscle, short-term); otoliths provide increased resolution (lifetime), but lethal
Genetic markers	SNP parentage assignment	Retrospective study; useful for untaggable fish	Requires full life-cycle movement models (including larval advection)
Parasite markers	Tetracotyle metacercarian parasite bio-tag data	Retrospective study; provides time-frame of exposure & location; useful for untaggable fish; cheaper than tags; elucidates behavioral differences	Complete knowledge of parasite biology & ecology required; taxonomic identification required; age of fish often required

Occurrence/density in predator food- habits samples	Sea birds as biological samplers for forage fish densities	Available opportunistically due to protected-species sampling programs; Predators are efficient samples of some prey species	Implicitly depends upon understanding of predator selection and functional responses; likely provides a noisy measurement of target species densities
---	--	---	--

896

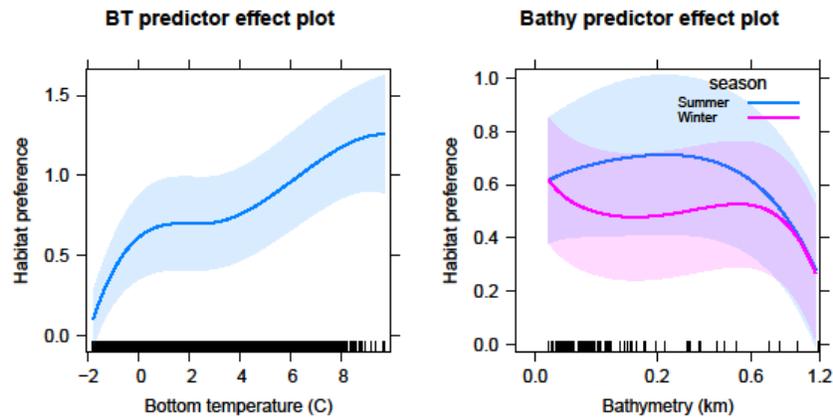
897

898 Fig. 1: Simplified illustration of diffusion-taxis movement model along a 1-dimensional spatial
899 domain defining habitat, $0 \leq x \leq 2\pi$, binned evenly into 25 grid cells $g \in \{1, 2, \dots, 25\}$, where
900 depth $x_1(g) = x/2$, temperature $x_2(g) = 3 + 2\cos(x)$, habitat preference $h(g) = x_1(g) +$
901 $x_2(g)$ (see panel A), diffusion rate $\tau = 4$, and assuming a reflective boundary (i.e., individuals
902 do not emigrate). The instantaneous movement rate matrix \mathbf{M}^* from each to every other grid cell
903 is the sum of the diffusion rate matrix \mathbf{D} (panel B) and the taxis rate matrix \mathbf{T} (panel C), where
904 $d(g_1, g_2) = t(g_1, g_2) = 0$ for any two grid cells that are not adjacent $g_1 \neq g_2 \pm 1$, otherwise
905 $d(g_1, g_2) = \tau$ and $t(g_1, g_2) = h(g_1) - h(g_2)$, and where the diagonal $d(g, g)$ and $t(g, g)$ is
906 defined such that columns sum to one (i.e., abundance is conserved during movement). For an
907 individual starting at $t = 0$ in grid cell $g = 3$ (panel D), $g = 13$ (panel E) or $g = 23$ (panel F),
908 we then show movement probabilities after one interval, $\text{matexp}(\mathbf{M}^*)$, three intervals,
909 $\text{matexp}(3\mathbf{M}^*)$, ten intervals, $\text{matexp}(10\mathbf{M}^*)$, or its limit after a long time has passed (i.e., long-
910 term habitat utilization, calculated as the dominant eigenvector of $\text{matexp}(\mathbf{M}^*)$). We note that
911 long-term habitat utilization is identical regardless of the initial location, and it resembles (but is
912 not identical to) the habitat preference function.



913

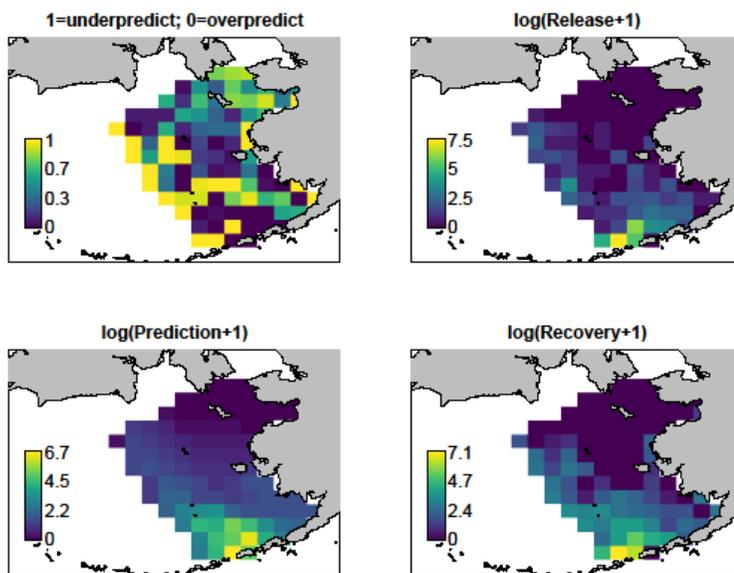
914 Fig. 2: Habitat preference (y-axis) for covariate values (x-axis) for each modeled covariate,
915 showing mean response (line) and 95% confidence interval (shaded area), visualized using the R
916 package *effects* (Fox & Weisberg, 2018)



917

918

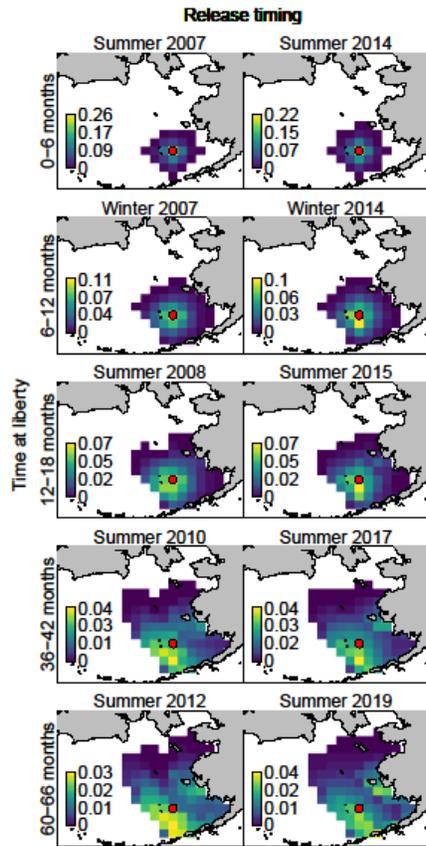
919 Fig. 3: Map of Probability Integral Transform (PIT) residuals for the recovery location of all
920 tags (top-left panel), the release location for tags (top-right column), the predicted recovery
921 location (bottom-left panel), and the observed recovery location (bottom-right panel), when
922 summing all tag releases, recoveries, and quantile residuals across all tags, seasons, and years.
923 The residuals are provided to illustrate any systematic spatial patterns in the sign of residuals.



924

925

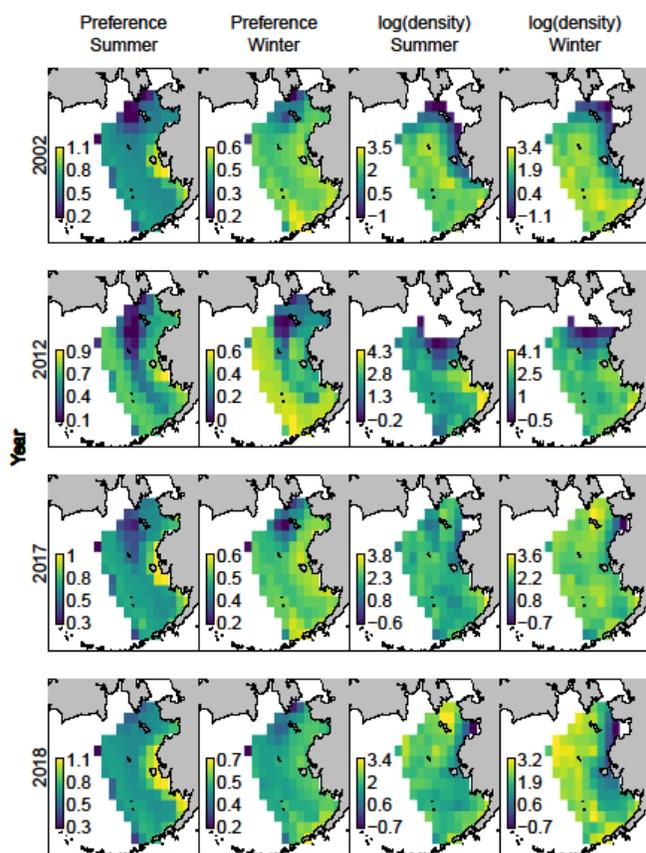
926 Fig. 4: Maps of movement probabilities (see color bar in each panel, and with probabilities <1%
927 of maximum plotted as white) given a southern release location in summer 2007 during a cold
928 stanza (left column) and summer 2014 during a warm stanza (right column)



929

930

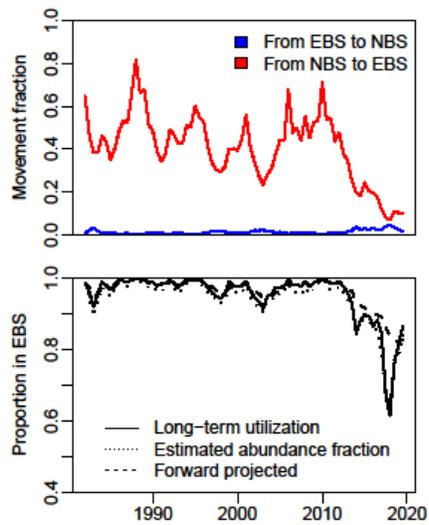
931 Fig. 5: Maps of habitat preference in summer (1st column) and winter (2nd column), as well as
 932 predicted log-abundance in summer (3rd column) and winter (4th column) for selected years
 933 (rows) highlighting a warm year (2002), cold year (2012) and then recent warm years (2017-
 934 2018).



935

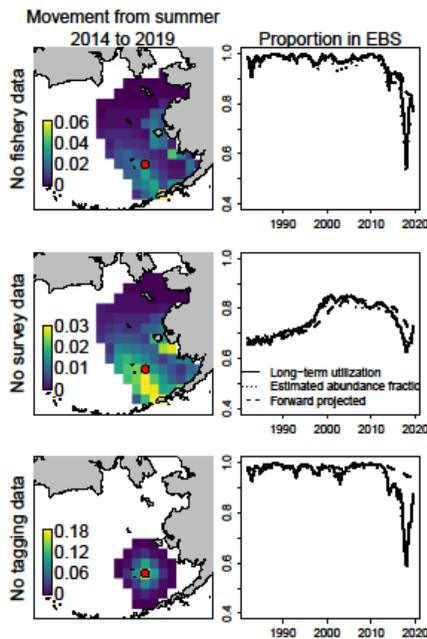
936

937 Fig. 6: Estimated summer-to-summer movement fraction (y-axis) from NBS to EBS (red line)
938 and EBS to NBS (blue line) in each year (x-axis)



939
940

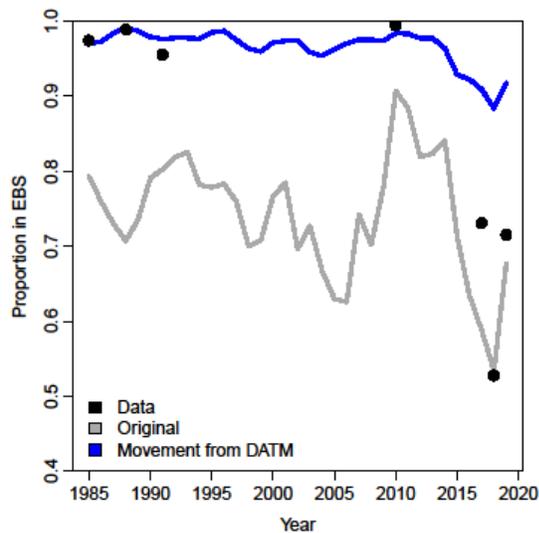
941 Fig. 7 – Sensitivity analysis exploring the impact of excluding each major data type including
942 fishery catch-and-effort (top row), survey data (middle row), or tagging data (bottom row),
943 showing movement (see Fig. 4 caption for legend details) from summer 2014 to summer 2019
944 (left column; see Fig. 4 bottom-right panel for plot for base model) and proportion of Bering Sea
945 abundance in the eastern Bering Sea in each year (right column; see Fig. 6 bottom panel for plot
946 for base model)



947

948

949 Fig. 8 – Illustrating survey and stock-assessment model estimates of the proportion of Bering Sea
950 Pacific cod summary abundance (defined as abundance for ages ≥ 1) occurring in the eastern
951 Bering Sea relative to the total summary abundance for northern and eastern Bering Sea strata
952 (y-axis) in each year 1990-2019 (x-axis) with legend in bottom-left corner (black bullets: survey
953 abundance converted to a proportion; grey line: original two-box age-structured stock assessment
954 model estimate; blue line: assessment model estimate after including the movement fractions
955 estimated by the DAT model). Note that the blue line deviates from the “forward projected”
956 estimates from the movement model (Fig. 6 bottom panel) because the stock assessment model
957 includes additional demographic structure (age-structure) and data types (age and length
958 composition samples).



959